

The INTSUM, RULEGEN, RULEMOD sequence works well for characterizing a set of molecules. If the set is representative of the whole class then the rules are general, otherwise they may be severely limited in scope. We want to give the program a sense of its limitations so that we can impart it to the chemists.

In particular, we want the program to request new data that will, in some sense, put the inferred rules to the test. And we want to be able to interpret the results of those experiments as indications that particular rules are weak, overly general, etc.

#### 2.8.6 Refinement of Parameters

We want to refine the parameters controlling the operation of the program so that there is a good "default" mode of operation. Because different persons like to focus on different aspects of problems, we want to leave open the option of a person setting the parameters for a very specific purpose. In order to do that, we need to characterize them in a way that chemists readily understand them.

#### 2.8.7 Interactive Programs

Part of the above concerns giving the chemist immediate control of the program. INTSUM and RULEGEN are very interactive now. In order to increase the use of RULEMOD (so we can get the feedback we need as well as to make the programs useful) we are about to start on making it interactive and helpful. Among other things we want to be able to answer questions on-line rather than having to look through hard-copy records of the programs' results.

#### 2.8.8 Merging Two Sets of Rules

The RULEMOD program can merge rules from the output of a single run of RULEGEN. We want to extend this capability so that we will be able to merge different sets of RULEGEN results, first from very similar molecules then from different classes of molecules.

### 3 PART 3: APPLICATIONS TO BIOMEDICAL STRUCTURE ELUCIDATION PROBLEMS

#### 3.1 Introduction

In our grant proposal we discussed the application of the instrumentation and computer programs described above to the

study of molecular structure problems in a variety of biomedical applications areas. This is our primary research area, and we discussed specific classes of problems and compounds for investigation. We also made it quite clear that our facilities would be made available to wider community of collaborators/users as our resources permitted. Both categories of application, i.e., within our own group, and with an outside group, are described in some detail below.

We have taken several steps toward encouraging a broad community of potential users to call on our facilities. For example, Appendix 2 contains the memorandum which was sent to local persons who had indicated their potential need for our facilities as described in our proposal. A questionnaire sent to members of the American Society for Mass Spectrometry, Committee III on Computer Applications, resulted in about 55 persons (Appendix 3) indicating a desire to know more about access to our programs. Appendix 4 contains the note which was sent to these persons. The same note has been sent to several other persons whom we know from personal contacts might be interested. Because of the nature of their investigations, many of these people receive NIH support.

The availability of SUMEX as a mechanism for resource sharing has made it possible for us to extend access to our programs to a number of people. Without SUMEX, this access would be impossible, and most of our programs (those which are not easily exportable) could be used only by ourselves.

We have coupled the above efforts with publications (see references 45-49) mentioning explicitly the availability of our programs. Through these efforts, together with talks and informal discussions we are slowly building a local and remote user community.

### 3.2 Applications by Professor Djerassi's Research Group

Our existing grants, outlined below, mesh well with our instrumentation and program development under the present award. Under NIH Grant GM06840 we have been studying natural products from marine sources with major emphasis on terpenoids and sterols. For this work we have been dependent on the use of our 711 instrument for high resolution mass spectrometry which we require for the identification of all new compounds, many of which are present in only very small quantities. We are particularly anxious to have access to GC coupled with a high resolution mass spectrometer because we hope to be able to screen large numbers of marine animals for their sterol content using this technique. In fact one of Prof. Djerassi's graduate students, Mr. R. Carlson, is currently working on a computer program which will automatically "reject" (i.e., detect, note, and not consider further) known sterols and point toward the presence of unknown ones which will then be the subject of further chemical work. In the context of terpenoids we have used

the INTSUM program with great effect in the structure elucidation of a group of new sesquiterpenes based on the novel skeleton I. A typical example of the sesquiterpenes is II, and the PLANNER, INTSUM and CONGEN programs coupled with high resolution mass spectrometry has been very helpful in elucidating its structure and that of other oxygenated analogs, and in understanding better the mass spectral behavior of these compounds.

Partly under NIH Grant No. GM06840 and partly under Grant No. AM04257 we have been working on elucidating the course of the mass spectrometric fragmentation of steroids and terpenoids through the use of labeled analogs. This information is of fundamental importance in order to apply it subsequently to the structure elucidation of new compounds, and here we have frequently needed access to the 711 instrument because of its "superhigh resolution" capability which was needed for distinguishing between compounds containing only C and H and compounds containing C, H and D. Furthermore, we have needed this instrument for metastable defocusing work and in fact hope very much that this will be eventually automated since this would represent a major simplification for much of our mass spectrometric research work.

Our current and proposed work with our programs for computer-assisted structure elucidation is discussed below under headings consisting of program names, which correspond to the programs discussed in Part 2. Much of the effort in application of a program(s) to the mass spectral data implicitly assumes that the data are available. In fact, without the current and future instrumentation effort discussed in Part 1, these program applications would not be feasible.

### 3.2.1 CLEANUP

The spectral cleanup program, written for ourselves and our collaborators in the Dept. of Genetics, Stanford Hospital (see Local/Stanford Community, below) will be included as part of the GC/LRMS system on the MAT-711 spectrometer. The essential nature of this program to treatment of the data prior to any more detailed examination was discussed previously. Although it is currently being tested, and now used routinely, for LRMS data from a quadrupole mass spectrometer, it is insensitive to the source of these data.

### 3.2.2 MOLION

This program is currently in routine use as an adjunct to LRMS data analysis subsequent to CLEANUP. It is incorporated as part of PLANNER as one way to detect molecular ions prior to analysis of the spectrum in terms of structure. Like CLEANUP, this is essentially a utility program, but plays a crucial role in applications of the other programs to structure elucidation problems.

### 3.2.3 PLANNER

PLANNER is currently being used to test the validity and generality of new mass spectrometry rules derived from INTSUM for the compound classes discussed under INTSUM, for example, the keto-androstanes and capnellanes. Such tests are important to ensure that existing rules can be safely extended to new (perhaps unknown) compounds in the same class.

A planned, major application of PLANNER was mentioned briefly above. The screening of marine sterols will use both GC/LRMS and GC/HRMS. PLANNER will be utilized to examine each spectrum and perform the task it does best, deciding where new substituents are most likely to be found about the steroid (cholestane) skeleton. The interactive nature of PLANNER simplifies the task of providing rules of mass spectrometry and constraints to the program. The rules in the case are the known mechanisms of fragmentation of various classes of sterols. In this way, known substances can be readily identified, even in the presence of other components, and suggested structures obtained for unknown compounds.

#### 3.2.4 INTSUM

As a means of extending the rules of fragmentation in mass spectrometry, several classes of compounds are under study as we attempt to determine characteristic modes of fragmentation. The following is a brief description of each such class and the current status of our research:

1. Pregnanes: Pregnanes related to the progesterone skeleton have been analyzed in some detail in collaboration with Dr. S. Hammerum, (University of Copenhagen, Denmark). One manuscript on this work has been accepted for publication in TETRAHEDRON [51]. One manuscript has been submitted to STEROIDS [52].
2. Androstanes: Keto-substituted analogs of the skeleton of the important steroidal hydrocarbon, androstane, are being studied in collaboration with Dr. Roy Gritter (an IBM scientist who spent his sabbatical leave in our laboratory learning more about mass spectrometry). This study is important to our understanding of the mass spectral behavior of complex, polycyclic systems. It is providing a model for the use of RULEGEN (see below). We are currently analyzing existing data and must acquire additional mass spectral data on new compounds.
3. Macrolide Antibiotics: We are in the middle of an interesting investigation of the fragmentation of macrocyclic antibiotics related to methmycin and neo-methmycin. INTSUM is proving very valuable in determining the regularities in the fragmentation behavior of these polyfunctional and polyheteroatomic compounds.
4. Phytoecdysones: These analogs of insect moulting hormones present difficult analytical problems in analysis of their

structures. We have been analyzing the mass spectra of several of these compounds to determine the feasibility of using PLANNER as a means of screening mixtures of these compounds for new structural types. Unfortunately, their tendency to dehydrate under even the most careful experimental conditions within the mass spectrometer means that spectra are obtained which do not contain as much structural information as we would like. We have confirmed, corrected and extended previous worker's results on diagnostic fragment ions and our investigations are continuing.

5. Insect Juvenile Hormones: In collaboration with Dr. Loren Dunham, Zoecon Corp., we are investigating regularities in the fragmentation behavior of the juvenile hormones. Previous work on the mass spectra of these compounds was carried out only at low resolving powers. So the simple determination of the HRMS will allow re-examination of past results. We have currently a set of representative compounds and their spectra. INTSUM work is now beginning.

#### 3.2.5 RULEGEN

As described above, RULEGEN can be used to assist in discovery of mass spectrometry fragmentation rules which depend on substructural features of molecules. Thus, it can be used for classes of compounds where the fragmentation does not depend on the basic skeleton, but on the positions of substitution. The keto-androstanes represent a case in point, as the fragmentations of these compounds are complex functions of the position of the keto group and the androstane skeleton itself. We are currently using RULEGEN on both "simple" classes of compounds and the estrogenic steroids as well as the androstanes.

#### 3.2.6 CONGEN

We are currently engaged in efforts to explore the utility of CONGEN to a variety of structure elucidation problems. The current areas of application are summarized below, together with progress to date.

- 1) Chlorinated Hydrocarbons: As an important class of a more general problem area in chemical structure analysis, the isomerism various types of chlorinated hydrocarbons has been investigated. The general structural problem is identification of possible substitution isomers about a given skeleton. The chlorinated hydrocarbons are, in addition, an important environmental and health problem. Using the labelling algorithm [41] we have constructed possible isomers of several classes of chloro-carbons [46].
- 2) Vertex-Graphs and Ring Systems: CONGEN has been used to explore possible ring systems in organic chemistry [47]. These features of the program have been utilized in several of the studies described below.

- 3) Ion Structures: CONGEN has been used to construct possible ion structures under a variety of constraints in support of studies on the structures of ions in the mass spectrometer. These studies are crucial to a deeper understanding of molecular fragmentation. The programs results (manuscript in preparation) are used to ensure that no plausible alternatives have been overlooked during efforts to characterize the structures.
- 4) Terpenoid Systems: We are using CONGEN to explore questions of the scope of terpenoid isomerism. We would like to determine some criteria which might allow us to say something about why only certain structural types are found in nature, to the exclusion of many possibilities which are very similar in structure.
- 5) General Structure Elucidation Problems: We are currently using CONGEN in two modes in connection with the structures of new compounds which have arisen in our recent research on marine organisms. The first mode has been to test several cases for which a structure had been proposed, to ensure that no other reasonable candidates had been overlooked. The second mode is in suggestion of structural possibilities for as yet unknown compounds, as we attempt to narrow the problem further by examination of candidate structures and design of experiments to differentiate among them.
- 6) Scope of Structural Isomerism: We are investigating the philosophical and pedagogical aspects of the scope of structural isomerism. This investigation is important to our program design and strategy as we identify the ways persons consider and reject whole categories of structural possibilities. The important artificial intelligence aspects of CONGEN lie in its ability to reason about molecular structure using efficient problem-solving strategies.

### 3.3 Applications by Other Members of the Stanford Chemistry Dept.

- 1) Prof. Mosher: We have used CONGEN to suggest structural possibilities for a naturally occurring analog of the fish poison tetrodotoxin. This structure is still under investigation.
- 2) Prof. Hahn (on sabbatical leave at Stanford from Syracuse University): We have used CONGEN to explore possible structures for unknown products of a photochemical reaction. These results have led him to begin a new set of experiments (specifically, CMR) to greatly restrict the possibilities.
- 3) Prof. Johnson: In his wide-ranging syntheses of steroid hormones and other steroids of biological interest, he has studied reactions involving stereo-specific cyclization. We are investigating use of CONGEN for structural analysis under

constraints imposed by synthetic cyclization experiments. For example, a previously investigated compound was found to have two structural possibilities. The new possibility could not be differentiated from the assigned structure based on available data.

- 4) Prof. Collman: We have utilized our mass spectrometry facilities to analyze samples in support of his work on oxygen binding to porphyrins (hemoglobin models).
- 5) Prof. Van Tamelen: We have provided mass spectrometry support (HRMS) to assist in the characterization of several compounds related to his work on terpenoid cyclizations.

#### 3.4 Applications by Other Stanford University Scientists

- 1) Genetics Research Center (GRC) Stanford Hospital: One of our strongest collaborations because of their requirements for additional automation in data reduction and analysis. Their screening program for metabolites characteristic of diseases of genetic origin uses GC/LRMS as the primary source of data. The CLEANUP and MOLION programs were written at least in part to assist the GRC in more systematic approaches to their data. We are currently using CONGEN to assist in determination of structures of unknowns for which mass spectrometric and chemical data are available. Our GC/HRMS facilities will also be utilized for problems which require determination of empirical formulas for ions in spectra of unknown compounds.
- 2) Stanford Pharmacy: We have had several requests for assistance from the Pharmacy of Stanford Hospital (Director: Dr. Hiram Serra). These have variously involved analyzing the stability and purity of pharmaceutical preparations, in particular: a. the impurity of stock preparations b. the stability of nitroglycerine tablets to heat; c. the stability over several months of methyl-dopa, prednisone and banthine when these compounds were formulated into syrups.
- 3) Drug Assay Laboratory Department of Pharmacology, Stanford University: Research personnel from this laboratory (Director: Sumner M. Kalman) have requested mass spectra on various derivatives of digoxin using both high and low resolution data.
- 4) Department of Psychiatry, Stanford University: The research group headed by Dr. J. Barchas has used low resolution mass spectral data for the purpose of structure elucidation of a basic compound of interest to their research program.
- 5) Department of Anesthesia, Stanford University: The DENDRAL group was asked by Dr. J. Trudell to help him in the identification of a urinary metabolite isolated after the administration of an anesthetic. This work involved high

resolution mass spectrometry of fractions isolated by Dr. Trudell.

- 6) Department of Psychiatry, Palo Alto Veterans Hospital: In this work we analyzed samples by GC/MS given to us by Dr. S. Kanter who works with Dr. Hollister. They were interested in detecting cannabinal, delta-9-tetrahydrocannabinol and an unknown (molecular weight 312) from urine extracts of subjects who had smoked marijuana. This involved running standards of cannabinal and its delta-9-tetrahydro analog through the GC/MS. We were unable to identify these compounds by mass spectrometry as being present in urine. In a subsequent meeting we learned that their concentration was less than 20 nanogram (per GC/MS injection) which is below the limits of sample flow for the recording of reproducible mass spectra. Dr. Kanter is working on the problem of isolating sufficient material for GC/MS and we expect to continue this project in year II of the current grant.
- 7) Prof. McCarty - Civil Engineering: Prof. McCarty is involved in a project to monitor water quality of effluents from tertiary sewage plants. This project includes significant efforts at characterization of the organic content of the water in various phases of its treatment to determine the efficiency of removal of various materials and to identify unknown organic compounds. We have agreed to provide instrumental and computer program support where necessary to assist him in characterization of these samples.

### 3.5 Applications by Non-Stanford Scientists

As an additional component of the resource sharing aspects of research, we have, as resources allow, extended the use of our facilities to a group of users remote from the local Stanford community. We have divided these users into two categories, those for whom we have provided mass spectrometry support and those who represent users of DENDRAL programs and collaborators on program development via the SUMEX resources.

#### A. Users of Mass Spectrometry Facilities

- 1) Professor O. O. Orazi, La Plata, Argentina: During the past year we have supplied Dr. Orazi with three low resolution mass spectra. We will be providing HRMS data for him in year II of our grant.
- 2) Professor T. Nakano, Caracas, Venezuela: Dr. Nakano sent one sample of an unknown alkaloid for high resolution mass spectrometry. We were able to show that his low resolution mass spectrum was 2 amu from the true molecular ion and after recording a low resolution mass spectrum his alkaloid was identified as a known compound.



- 3) Dr. Steen Hammerum, Copenhagen, Denmark: Dr. Hammerum requested our assistance in running ultra high resolution mass measurements on several ions in the mass spectra of compounds he had specifically labelled with  $^{13}\text{C}$ .

#### B. Users/Collaborators of/with DENDRAL Programs on SUMEX

Below, in alphabetical order, we list those persons who have a) expressed interest in use of our programs and have been sent instructions in how to gain access to SUMEX and our programs. In many cases these persons have received more detailed information in the form of demonstrations in person or remotely using the LINK facilities of SUMEX, and b) persons who have acted as collaborators in development of parts of one or more of our programs. (Some persons fall in both categories.) Each prospective user generally receives the following packet of material:

i) An Introduction to SUMEX-AIM - a simplified guide to the SUMEX system for those who do not need the complete TENEX operating system manual.

ii) Network information - TYMNET or ARPANET instructions on how to connect to the network and use the network to connect to SUMEX.

iii) Account and password information.

iv) Documentation for the specific programs the user wishes to access.

v) Explanation of the SUMEX RECORD facility which allows us to examine a user's complete terminal session when he has encountered trouble.

Because we have just begun encouraging a significant community of persons to try our programs, we do not yet have a good idea of which persons will continue as serious users. But we have at least provided the opportunity for persons to gain access to our programs, try them and determine how they might (or might not) fit into their own research problems. The term "exploratory" refers precisely to this category of persons who are now engaged in this kind of evaluation. The program names after each person's activity refer to their current major interest. In some cases, we do not actually know the specific problems which are being explored.

1. Dr. A.L. Burlingame (U.C. Berkeley) - Exploratory - all DENDRAL programs.
2. Prof. E.J. Corey (Harvard) - Exploratory, collaboration on programming strategies, CONGEN.
3. Dr. L. Dunham (Zoecon) - Exploratory - Structure determination - CONGEN.

4. Dr. H.M. Fales (NIH) - Exploratory - all DENDRAL programs.
5. R. Feldmann (NIH) - Collaborative development of programs (structure input and drawing routines).
6. Prof. D.L. Fishel (Kent State) - Considering access to SUMEX - has the program descriptions.
7. Prof. M.J. Goldstein (Cornell) - We have provided CONGEN results to him for a difficult structure problem.
8. Dr. N.A.B. Gray (Cambridge) - Collaborating on strategies for computer-assisted structure elucidation programs. He is working on spectral data interpretation.
9. Dr. P. Gund (Merck, Sharpe & Dohme) - Arranging an on-line demonstration for exploratory purposes - CONGEN.
10. Dr. J. Karliner (Ciba-Geigy) - Using CONGEN on structure elucidation problems.
11. Dr. S. Heller (Environmental Protection Agency) - Collaboration on mass spectral library development.
12. Dr. P. Jurs (Penn. State) - Collaboration on structure analysis and building of chemical structure models.
13. Dr. B. Kowalski (Univ. of Washington) - Has approached us for use of SUMEX in pattern recognition work.
14. Dr. D. Lefkowitz (Univ. of Penn.) - Exploratory - interest in DRAW portion of CONGEN for NCI chemical information system.
15. Dr. S. Markey (NIH) - Exploratory - all DENDRAL programs.
16. Dr. F. McLafferty (Cornell) - Exploratory - all DENDRAL Programs.
17. Dr. R. Milberg (National Center for Tox. Res.) - Exploratory - CONGEN.
18. Dr. D. Poulter (Univ. of Utah) - Exploratory - use of CONGEN in structure determination problems, especially terpenoids.
19. Dr. K. Rinehart (Univ. of Illinois) - Exploratory - all DENDRAL programs.
20. Dr. P. Roller (National Cancer Institute) - Exploratory - all DENDRAL programs.
21. Dr. R. Rosen (FMC Corp.) - Exploratory - all DENDRAL programs.
22. Dr. G. Szonyi (Polaroid Corp.) - Interest in CONGEN, exploratory phase beginning.

23. Dr. W.T. Wipke (Princeton) - Exploratory - CONGEN  
collaboration on structure model building and methods for  
stereochemical representation of chemical structure.

#### 4     BIBLIOGRAPHY

See Section II-D, SUMMARY OF PUBLICATIONS, for a listing of publications by this project.

Figure 1. Mass Spectrometry  
Data Acquisition Hardware  
Configuration

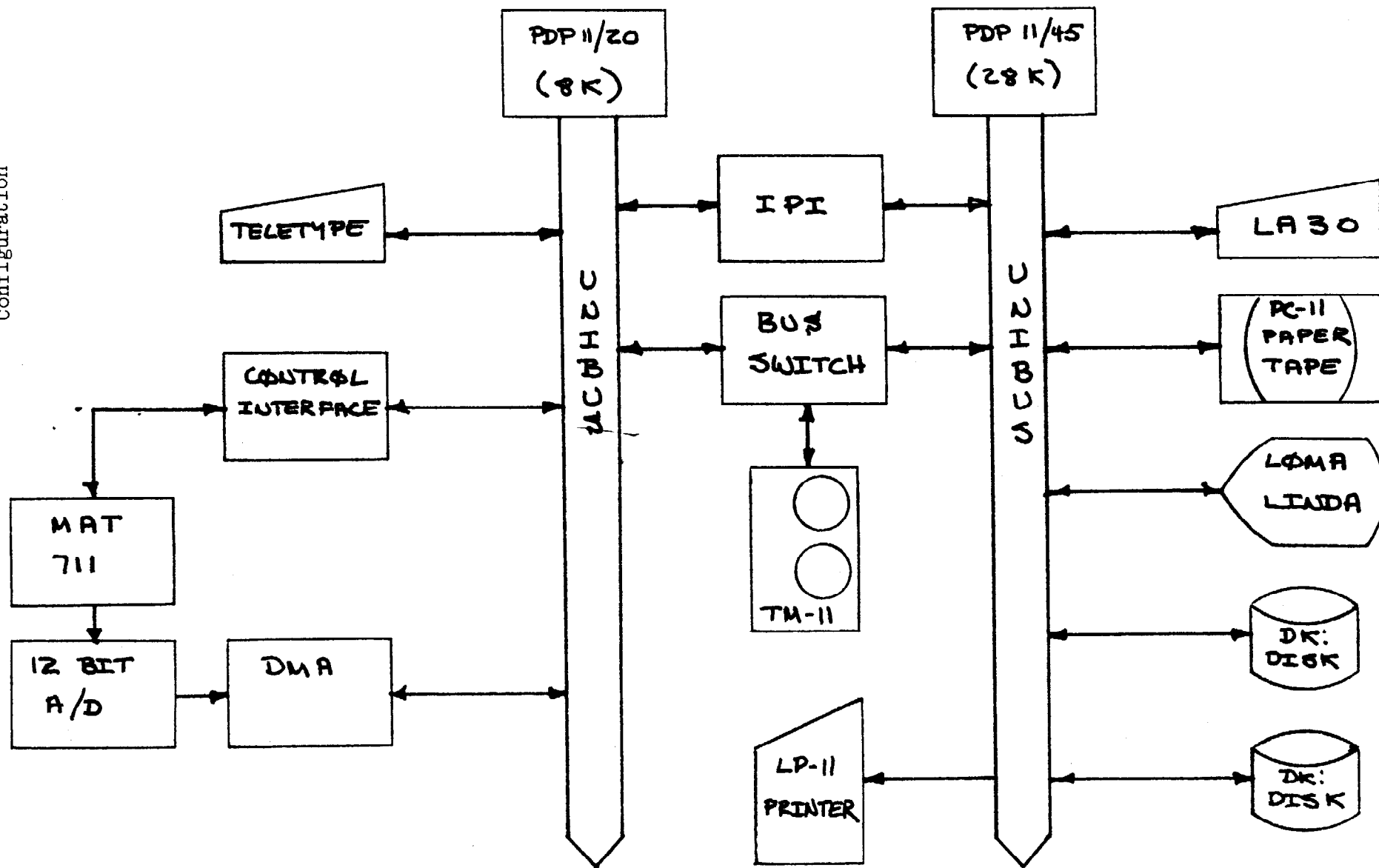
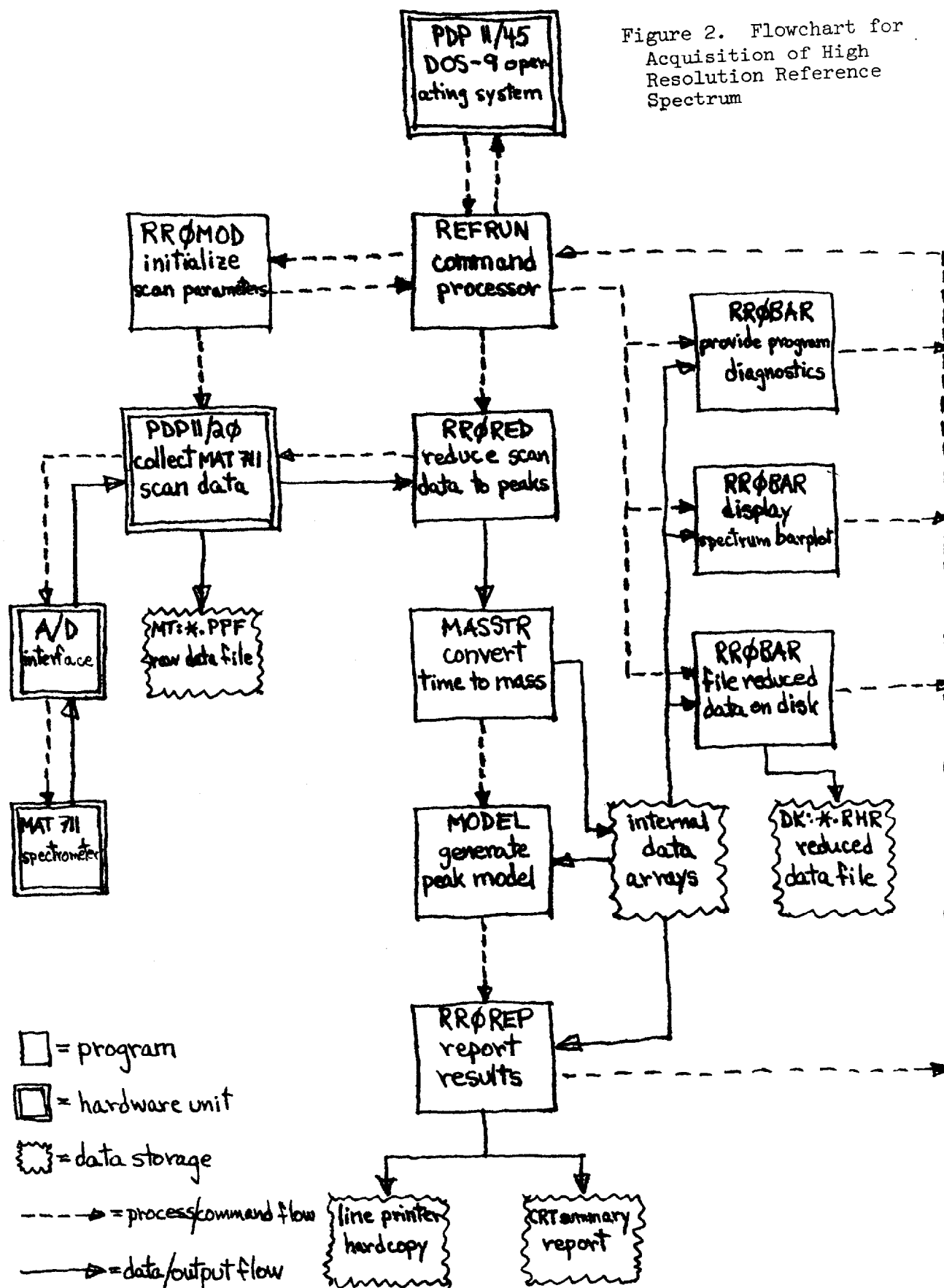
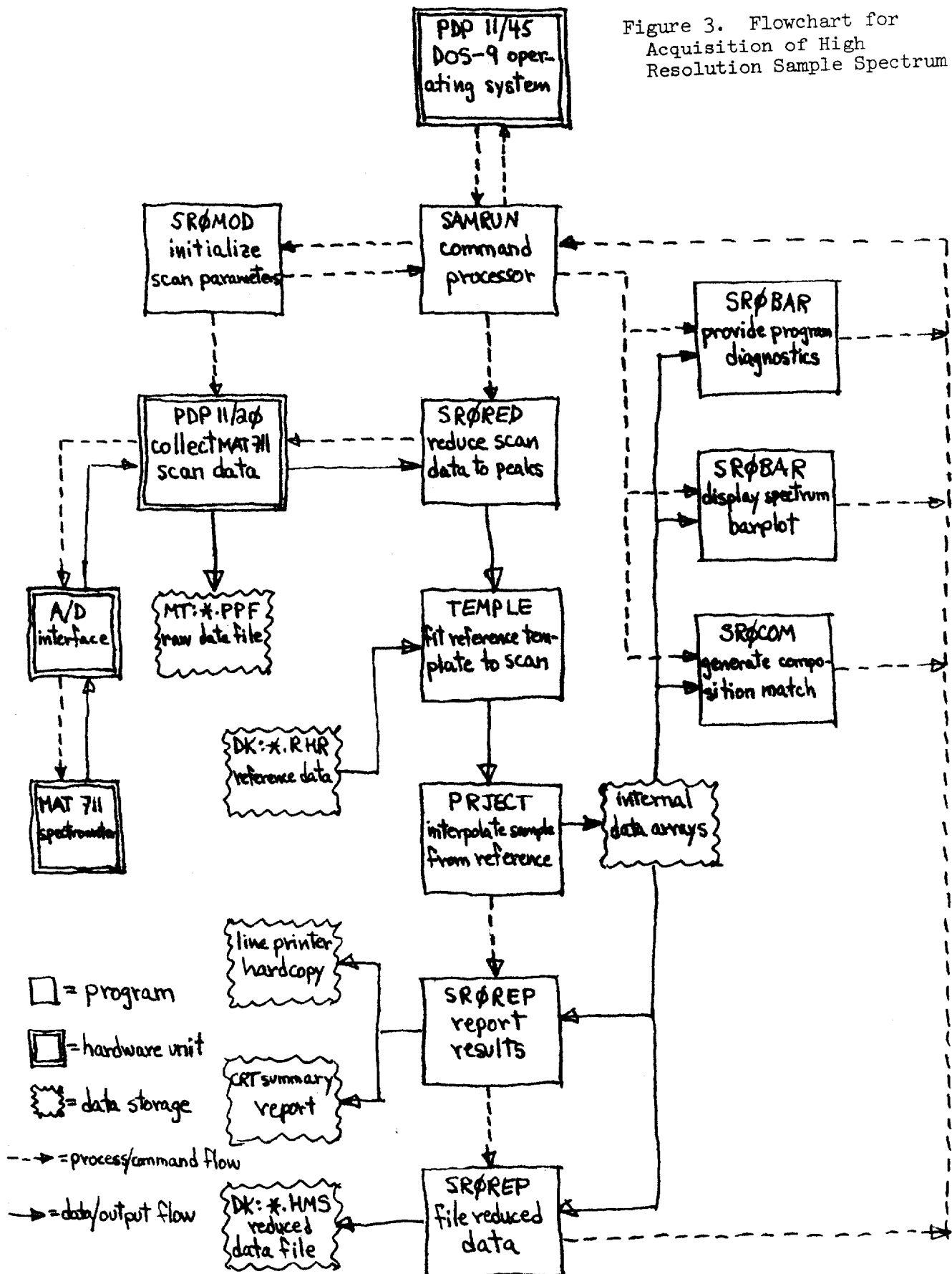


Figure 2. Flowchart for Acquisition of High Resolution Reference Spectrum



Flowchart for Acquisition of High Resolution Reference Spectrum

Figure 3. Flowchart for Acquisition of High Resolution Sample Spectrum



Flowchart for Acquisition of High Resolution Sample Spectrum

Appendix 1  
Typescript of Interactive Session with  
The DENDRAL PLANNER

Annotations are bracketed and prefixed with asterisks.

[\*\*\* 1. Start the Planner program.]

@WORK5  
INTERLISP-10 20-OCT-74 ...

Good morning, Bruce.

(WORK5.;1 . <SUBSYS>NLISP.SAV;1)

\_RUN]

[\*\*\* 2. After the program prompts for class name it asks if the structure and fragmentation rules for the class have been stored in a file from a previous session.]

CLASS NAME:CAPNELLANE  
GET CLASS PARAMETERS FROM A FILE (Y/N) ? Y  
GET CLASS FILE NAMED: <SMITH>CAPNELLANE.PLANTEST  
FILE CREATED 20-MAR-75 14:08:31  
&  
YOU HAVE JUST SET PARAMETERS FOR THE CLASS CAPNELLANE  
SKELETAL COMPOSITION: ((H . 26) (C . 15))

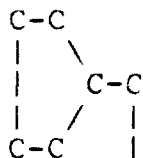
[\*\*\* 3. The data are now read in from a file and adjusted for isotopic contributions, with peaks below the noise threshold deleted.]

READ A NEW MASS SPECTRUM (Y/N) ?Y  
NAME OF SPECTRUM FILE: <SMITH>1CAPNELLANE.TEST  
HIGH RESOLUTION  
MASS RANGE: 40.0317 TO 234.1621 (82 PEAKS)  
TOTAL ION VALUE: 1440.652  
(PEAKS SMALLER THAN 7.203262 DELETED.)  
PRINT ? N

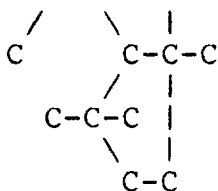
[\*\*\* 4. The program is ready to begin its analysis at this point. However, the user asks instead to see the structure and breaks for this class, to be sure that they were set correctly in the previous session.]

START PLANNING (Y/N)?N  
WANT HELP (Y/N)?N  
STRUC?

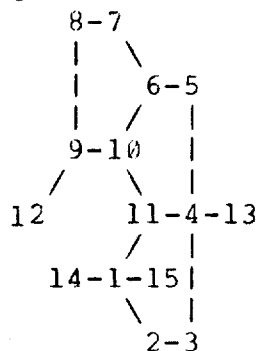
CAPNELLANE







CAPNELLANE



BREAKS?

-----  
 LABEL: 6H  
 BONDS: ((3 - 2) (1 - 11))  
 HTRANS: (-1 0)  
 -----

LABEL: 7H  
 BONDS: ((11 - 1) (3 - 4))  
 HTRANS: (-1 0 1 2)  
 -----

LABEL: 7L  
 BONDS: ((1 - 11) (3 - 4))  
 HTRANS: (-1)  
 -----

[\*\*\* 5. When the user examines the parameter CONTROLRULES, it has no effective value. So it is given a value in order to control the amount of evidence used by the program to build structures.]

CONTROLRULES?  
 CONTROLRULES= NIL  
 CONTROLRULES  
 USE ONLY THE STRONGEST EVIDENCE FOR SOME BREAKS (Y/N)?Y  
 APPLICABLE BREAKS:6H  
 APPLY ON PASSES:?

ENTER EITHER A LIST OF INTEGERS OR THE WORD ALL. THE INTEGERS INDICATE THE PASS NUMBERS (I.E. THE NUMBERS OF THE ATTEMPTS) OF THE STRUCTURE BUILDING PGM FOR WHICH THIS CRITERION IS APPLIED TO THESE BREAKS. E.G. PASS 1 BUT NOT AFTERWARD, PASSES 1 AND 2 BUT NOT AFTERWARD, OR ALL PASSES.  
 PASSES:1  
 USE EVIDENCE THRESHOLD FOR SOME BREAKS (Y/N)?Y  
 APPLICABLE BREAKS:7H 7L  
 APPLY ON PASSES:1  
 PERCENT OF MAXIMUM INTENSITY (DEFAULT = 33): ?

WHAT THRESHOLD DO YOU WANT TO USE AS A CUTOFF -- GIVE A NUMBER THAT  
WILL BE USED AS PERCENT OF MAXIMUM INTENSITY TO THROW AWAY (RELATIVELY)  
SMALL INTENSITY PIECES OF EVIDENCE.  
PERCENT: 33

[\*\*\* 6. Now we are ready to begin analyzing the mass spectrum.]

# PLAN

[\*\*\* 6a. The program first infers that the molecule contains  
15 carbons, 22 hydrogens, and 2 oxygens.]

(COMPUTING MOLECULAR ION(S))

MOLECULAR IONS

(234.1609 100 (C . 15) (H . 22) (O . 2))

[\*\*\* 6b. The program now looks for evidence (peaks in the mass  
spectrum) corresponding to possible combinations  
of residual atoms and double bonds in each fragment.]

(STARTING ANALYSIS PART)

BREAK : SUBSTITUENTS ON CHARGED FRAGMENT : EVIDENCE (M/E)

6H	((DOT . 4) (C . 0) (O . 1))	161.0975
	((DOT . 2) (C . 0) (O . 1))	163.1136
	((DOT . 4) (C . 0) (O . 2))	178.0993

---

7H	((DOT . 4) (C . 0))	133.1009	
	((DOT . 2) (C . 0))	133.1009	
	((DOT . 4) (C . 0) (O . 1))	150.1038	149.0951 148.088
147.0811	((DOT . 2) (C . 0) (O . 1))	150.1038	149.0951
	((DOT . 4) (C . 0) (O . 2))	163.0758	

---

7L	((DOT . 4) (C . 0))	65.03971
	((DOT . 2) (C . 0))	67.05513
	((DOT . 0) (C . 0))	69.07053
	((DOT . 0) (C . 0) (O . 1))	85.06461

---

[\*\*\* 6c. Structural descriptions are now put together in all  
plausible ways that are consistent with the substituent  
evidence just gathered.]

BEGIN SYNTHESIS OF MOLECULAR ION = 234.1609

STRUCTURE 1

((DOT . 4) (O . 1)) C4 C5 C6 C7 C8 C9 C10 C11 C12 C13)  
((O . 1)) C3)

EVIDENCE USED TO BUILD STRUCTURE:

(7H (DOT . 4) (O . 1))  
(6H (DOT . 4) (O . 2))  
(7L (O . 1))

-----  
DONE

Appendix 2  
Memorandum to Stanford Scientists

DATE: May 2, 1974

TO : Potential Stanford University Collaborators with the DENDRAL Project

FROM : Carl Djerassi, Professor of Chemistry

SUBJECT: Availability of Facilities

During the preparation (December, 1973) of our grant application entitled "Resource Related Research: Computers in Chemistry" your research group expressed interest in utilizing our facilities (see below) for assistance in solving structure elucidation problems related to health sciences. I now wish to notify you, in my role as P. I. for this grant, that it was funded as of May 1, 1974, for a three year period, substantially as requested.

I want to make a few general comments before describing the facilities which will become available during the course of this grant. Our primary goals, as spelled out in our grant application, deal with bringing state-of-the-art techniques in mass spectrometry and computer science to bear on solving problems of molecular structure. We are not funded to act as a general service facility; we have neither the time nor the personnel to function in this manner. We hope to operate in a collaborative manner with each of you to help decide questions concerning the specific instrumental and computer techniques which can be brought to bear on your problems.

#### Facilities

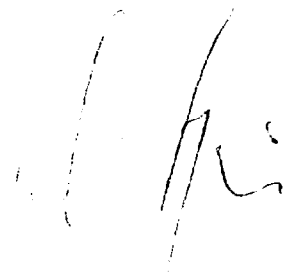
A) Mass Spectrometry. Our primary goal is to provide the capability for routine gas chromatography/high resolution mass spectrometry. We will shortly receive a PDP 11/45 computer system for the laboratory which will be programmed to carry out this task. At the present time we can provide gas chromatography/low resolution mass spectrometry and severely restricted (without charge) access to high resolution mass spectra of single compounds as our budget provides only minimal funds for supporting this work pending the completion of the 11/45 data system. These facilities are available at no cost to the user.

B) Computer-Assisted Structure Elucidation. We have available a number of programs designed for automatic analysis of mass spectral data and also for isomer generation and manipulation and display of chemical structures. We are designing and programming interactive systems which will allow users to answer problems concerning the identity of molecular structures based on a variety of spectroscopic data. These interactive systems are under development, but are always available in their present state for tolerant users. These programs run on the SUMEX (Prof. Lederberg, P. I.) PDP-10 at the Medical School. Details of local user access are being worked out, but a significant amount of computer time will be available free of charge for users of these programs.

For the present time, those of you who are interested in making use of these facilities should contact (Med. School) Dr. Alan Duffield (Ext. 7-5788 (temporary, soon to be 7-6389) or (Chemistry Dept.) Dr. Dennis Smith (Ext. 7-3144). If serious bottlenecks occur, either with respect to the mass spectrometry laboratory or SUMEX, I intend to make use of an advisory committee described in our grant application to help rectify the problem. It is my hope that judicious selection of problems along the lines of the collaboration I outlined above will not make this necessary.

CD:ab

cc: Prof. J. Lederberg, Genetics  
Prof. E. Feigenbaum, Computer Science

A handwritten signature, likely of the sender, is located in the upper right quadrant of the page. It consists of a large, stylized capital 'A' followed by a cursive 'L' and a small 'S'.

Appendix 3  
Questionnaire Sent to American Society for Mass Spectrometry,  
Committee III on Computer Applications

## DENDRAL LIST

<u>Name</u>	<u>Address</u>		
H.E. Lumpkin	Exxon Res. & Eng. Co. P.O. Box 4255	Baytown, Texas	77520
R.K. Mauldin	Celanese P.O. Box 1414	Charlotte, N.C.	28232
M.F. Zabielski	United Aircraft Res. Labs.	E. Hartford Conn.	06108
T.E. Edwards	Mail Stop ES 32 Marshall Space Flt. Ctr. E228/113 Exp. Station	Huntsville, Ala.	35812
C.N. McEwen	Dupont	Wilmington, Del.	19898
B. Tiffany	Savannah River Labs. Dupont	Aiken, S.C.	29801
J. Capellan	Ames Lab USAEC Iowa State	Ames, Iowa	50010
K.R. Thompson	Eng. Phys. Lab B-357 Dupont	Wilmington, Del.	19898
T. Krick	Biochem. Dept. U. of Minn.	St. Paul, Minn.	55101
J.R. Hass	NIEHS P.O. Box 12233	Research Triangle Park, N.C.	27709
F. Hileman	D. of Chem. U. of Utah	SLC, Utah	84112
M. Hoffman	D. of Chem Kansas State U	Manhattan, KS	66506
W.T. Rainey	P.O. Box Y Bldg. 9735 ORNL	Oakridge, TN	37830
M.E. Fitzgerald	Arco Chemical Co. R + E P.O. Box 85	Glenholden, PA	19036
W.L. Krudop	3316 Darryl Ln.	Modesto, Calif.	95350
B.A. Raby	325 N. Mathilda Ave.	Sunnyvale, CA	94086
W.C. Judd	Gen. Elect. 21800 Tungsten	Cleveland, OH	44117
R.W. King	U. of Florida Dept. of Chemistry	Gainesville, FL	32611
W. Killinger	W.R. Grace Co. Wash. Res. Ctr. 7379 RT32	Columbia, MD	21044
M.C. Manning	Continental Oil Co. R + D	Ponco City, Okla.	74601
J. Dillard	Chem. Dept. VA. Tech	Blacksburg, VA	24061
E.M. Chait	Dupont Instruments 1500 S. Shamrock Ave.	Monrovia, CA	91016



<u>Name</u>	<u>Address</u>		
R.T. Rosen	FMC Corp. P.O. Box 8	Princeton, N.J.	08540
J.C. Orr	Harvard Medical School 45 Shattuck St.	Boston, MA	02115
M. Levenberg	Dept. 482 Abbot Laboratories	North Chicago, Ill.	60064
J.M. McGuire	S.E. Environmental Res. Lab. College Station Rd.	Athens, GA	30601
B.S. Middleditch	Dept. Cell Biology Baylor Coll. Med.	Houston, Texas	77025
G.L. Kearns	The Kearns Group 58 Buckingham Dr.	Stamford, CT	06902
H. Fales	NIH	Bethesda, MO	20014
E. G. Perkins	104 Burnside Lab U of Illinois	Urbana, ILL.	61801
M.W. Karasek	Chem. Dept. U of Waterloo	Waterloo, Ontario	
J.M. Miller	D of Chem Brock University	St. Catharines, Ontario	
I.M. Campbell	RM 605 GSPH U of Pittsburgh, 130 Deste St.	Pittsburgh, PA	15261
K.A. Lincoln	NASA Ames Res. Ctr. Moffett Field	CA	94035
F. Falkner	Drug Metabolism Central Res. Pfizer, Inc.	Eastern Pt. Rd. Groton, Conn	06340
D.V. Bowen	1221 York Ave. New York, NY		10021
R.D. Grigsby	Dept. of Biochem and Biophys. Texas A + MU.	College Station, TX	77843
L.R. Yetter	IBM Corp. P.O. Box 417, 19 Barton Rd.	Apalachin, NY	13732
B.A. Allen Jr.	832 Sycamore Lake Jackson	TX	77566
H.S. Hertz	A105 Chemistry Nat'l. Bur. Stds.	Wash. D.C.	20234
R. J. Ayers	FDA 1560 E. Jefferson	Detroit, Mich.	48207
J.F. Paulson	A.F. Cambridge Res. Labs/LKB Hanscom AFB	Bedford, Mass.	01730
R.J. Weinkam	Dept. Pharm. Chem. U of Calif. S.F.	San Francisco, Calif.	94143
K.L. Rinehart	454 Roger Adams Lab U of Ill.	Urbana, Ill.	61801